EE524 Machine Learning Lab

Assignment 3

30 August 2022

- **Naive Bayes Theorem:**
  Rival teams in Cricket, India and Pakistan. India won the matches 65% of the time and Pakistan wins the remaining matches. Amongst the games won by India, only 30% of them come from playing in Pakistan's country stadiums. 75% of victories for Pakistan are obtained while playing at home. If India hosts the next match, which team is likely to emerge as the winner? Let X be the random variable that represents the team hosting the match, and Y be the random variable that represents the winner of the match. X and Y both are discrete random variables taking values in 0,1.

  Compute using Bayes theorem, write down functions for priors, likelihoods, evidence and posteriors and call them. Depending on the posteriors decide the winning team.

  Prior: P(Y)
  Evidence: P(X)
  Likelihood: P(X — Y)
  Posterior: P(Y — X) where X and Y take values in the set 0,1.

- **Naive Bayes Classification:**

  - Import the dataset provided in CSV format as a dataframe. Just looking at the data, predict which feature is the most discriminating for a person to be a defaulted borrower.
  - Convert the dataset into a numbered one as home_owner is a binary feature, marital_status is a categorical feature, annual_income has to be kept the same, defaulter is a class label which is again binary.
  - Consider this problem as a Gaussian Naive Bayes Classification. Using the same functions written in the first question, compute the likelihood functions, P(home_owner | defaulter) and P(marital_status | defaulter) using the frequentist approach. Compute the likelihood

function of P(annual_income | defaulter) through a Gaussian distribution with parameters as μ the sample mean of income and σ^2, the sample variance of annual income. The overall likelihood function P(X | Y) is the combination of likelihoods of 3 features (use conditional independence).

- Predict the status of a person as a loan defaulter for the following test data X test = (home_owner = no, marital status = married, annual income = 1,20,000). For this, compute the posteriors P(defaulter = yes | X test) and P(defaulter = no | X test). You can also consider the evidence P(X) or completely ignore it.